

ANALISIS DATA POLITIKUS MENGGUNAKAN NODE2VEC DAN ALGORITMA HARMONIC CENTRALITY

Ahmad Rifqy Syarwani¹, Erizkika Mochammad Arsy Rofi², Nur Aini Rakhmawati³

^{1,2,3} Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi
Sepuluh Nopember, Surabaya

¹rifqy.17052@mhs.its.ac.id, ²erizkika.17052@mhs.its.ac.id, ³nur.aini@is.its.ac.id

Abstrak

Indonesia adalah negara demokrasi yang memiliki partai politik untuk merepresentasikan suara rakyat. Politikus mendapatkan jabatan pemerintahan melalui afiliasi partai politik. Suara partai politik dapat mempengaruhi partai lain dan sistem politik secara keseluruhan. Karenanya, pengukuran pengaruh partai politik dilakukan untuk mengetahui dampak pada sistem pemerintahan. Menggunakan algoritma harmonic centrality untuk mengukur pengaruh suatu partai atau politikus yang difasilitasi oleh Node2Vec. Hasil analisa menunjukkan terdapat partai yang memiliki pengaruh terbesar pada pemerintahan. Namun graph memiliki density yang cukup rendah dikarenakan keterbatasan data.

Kata kunci : harmonic centrality, node2vec, politikus, analisa graf

Abstract

Indonesia is a democratic country with political party acting as people voices. Politician acquire governmental position through affiliation with a political party. The political party voices can influence other parties and the whole political system. Hence, measuring the influence of a political party is done to determine the impact in the government system. Using harmonic centrality algorithm to measure the influence of a political party or politician which facilitated by Node2Vec. Analysis results shows that there is a political party that has the most influence in the government. However the graph has quite low density due to data limitation.

Keywords: harmonic centrality, node2vec, politician, graph analysis

1. Pendahuluan

Indonesia merupakan negara dengan sistem politik demokrasi. Pada sistem tersebut terdapat partai politik yang merupakan kelompok terorganisir yang memiliki orientasi, nilai-nilai dan cita-cita yang sama dan dapat dipilih oleh rakyat pada pemilihan umum (Kadir, 2014). Politikus dapat memiliki jabatan/kursi dalam suatu pemerintahan dikarenakan dipilih oleh rakyat melalui afiliasi partai politik. Sebagai suatu sistem pemerintahan yang kompleks dan dinamis, sikap dan keputusan suatu partai politik dapat mempengaruhi satu sama lain dan dapat berdampak pada sistem politik secara langsung. Kekuatan suatu partai politik perlu diidentifikasi untuk mengetahui partai politik mana yang memiliki pengaruh besar terhadap partai lainnya. Partai dengan kekuatan besar memiliki pengaruh untuk membentuk koalisi yang akan diikuti oleh partai-partai lainnya.

Konsep centrality sering digunakan untuk mencari peran dan pengaruh dari suatu node dalam suatu jaringan. Harmonic centrality adalah konsep centrality dengan menghitung menggunakan harmonic mean (Marchiori & Latora, 2000). Penerapan konsep harmonic centrality digambarkan pertama kali oleh (Rochat, 2009), harmonic

centrality memiliki hubungan dekat dengan closeness centrality namun dengan keunggulan dapat menyelesaikan kemungkinan jarak tak terbatas dengan tepat.

Analisa sentralitas aktor dalam konteks politik dilakukan (Sugiarta, et al., 2018) untuk mengetahui aktor yang paling berpengaruh pada komunikasi politik. Penelitian yang dilakukan (Rochat, 2009) mengemukakan perbedaan antara closeness dengan harmonic centrality sehingga harmonic centrality dinilai lebih layak untuk diterapkan pada kasus dunia nyata. Penelitian lain dilakukan oleh (Boldi & Vigna, 2013) untuk mengoptimasi nilai centrality menggunakan HyperBall yang dapat mengurangi penggunaan memori secara drastis. Algoritma yang dikemukakan (Putman, et al., 2019) dapat mempercepat perhitungan centrality hingga 33 kali pada dataset directed dan weighted. (R., et al., 2019) Membuat model machine learning agar dapat memprediksi nilai centrality agar mengurangi beban komputasi.

Tujuan penelitian ini dilakukan adalah untuk mengetahui partai politik yang memiliki pengaruh besar terhadap sistem pemerintahan di Indonesia selama ini berdasarkan data dari DBpedia Indonesian. Selain itu dilakukan analisa kedekatan antara partai politik dan politikus melalui

pendekatan graf menggunakan algoritma harmonic centrality.

2. Dasar Teori

2.1. SPARQL

SPARQL merupakan serangkaian spesifikasi yang menyediakan bahasa dan protokol standar untuk melakukan *query* dan manipulasi data graf RDF. Sebagai bahasa *query* untuk RDF, SPARQL dapat mengambil data yang disimpan pada SPARQL service yang merupakan HTTP *service endpoint* yang dapat menjalankan SPARQL *queries*. *Query* yang dapat dijalankan terdiri dari *query* standar hingga kompleks seperti *union*, *optional query parts*, *filters* dan fitur baru yang ditambahkan di SPARQL 1.1 yaitu *value aggregation*, *path expression*, *nested queries*, dan sebagainya. Hasil dari SELECT *queries* di SPARQL terdiri dari pemetaan antara variabel dan RDF *terms* yang biasa divisualisasikan dalam bentuk tabel. Perihal untuk mengubahnya dalam bentuk yang dapat dengan mudah dibaca oleh mesin, SPARQL mendukung perubahan hasil format ke bentuk umum lainnya, yaitu Extensible Markup Language (XML), JavaScript Object Notation (JSON), Comma Separated Values (CSV), dan Tab Separated Values (TSV) (W3C, 2013).

2.2. Node2Vec

Node2Vec merupakan algoritma *semi-supervised* yang digunakan untuk pembelajaran fitur yang terukur pada sebuah jaringan. Pendekatan algoritma ini adalah mengembalikan representasi fitur yang memaksimalkan probabilitas dalam mempertahankan kedekatan jaringan *nodes* dalam sebuah d-dimensional ruang fitur. (Grover & Leskovec, 2016). Algoritma Node2Vec mempelajari secara kontinu representasi untuk *nodes* pada *undirected* dan *unweighted* graf. Hasil representasi dari *nodes* pada graf nantinya dapat digunakan untuk berbagai pekerjaan *machine learning* (Leskovec, 2016).

2.3. Algoritma Harmonic Centrality

Konsep harmonic centrality diusulkan oleh (Marchiori & Latora, 2000) dengan menggunakan harmonic mean untuk mendapatkan jarak konektivitas sebuah graph. Algoritma harmonic centrality dikembangkan untuk menyelesaikan permasalahan jarak antar nodes yang tidak terhubung, karenanya jarak bernilai ∞ . Harmonic centrality dapat menangani kasus ini dengan cara mensubstitusi nilai ∞ dengan ∞^{-1} , sehingga

ditetapkan algoritma harmonic centrality sebagai berikut (Boldi & Vigna, 2014).

$$\sum_{y \neq x} \frac{1}{d(y, x)} = \sum_{d(y, x) < \infty, y \neq x} \frac{1}{d(y, x)}. \quad (1)$$

3. Metodologi Penelitian

Gambar 1 menunjukkan diagram metodologi yang akan digunakan.



Gambar 1. Diagram Metodologi

3.1. Query SPARQL

Query dilakukan pada situs aplikasi SPARQL milik DBpedia Indonesian pada alamat id.dbpedia.org/sparql dengan sintaks yang ditunjukkan pada gambar 2.

```

SELECT DISTINCT ?Politikus, ?Partai WHERE {
  { ?p dbpedia-owl:knownFor
    <http://id.dbpedia.org/resource/Politikus> . }
  UNION { ?p dbpedia-owl:occupation
    <http://id.dbpedia.org/resource/Politikus> . }
  UNION { ?p dbpedia-owl:profession
    <http://id.dbpedia.org/resource/Politikus> . }
  ?p rdfs:label ?Politikus .
  ?p dbpedia-owl:party ?q .
  ?q rdfs:label ?Partai .
}
  
```

Gambar 2. Sintaks Query

Data yang didapat lalu dilakukan penyeragaman dan penghapusan data politikus pada partai dari luar negeri atau yang sudah tidak aktif beroperasi. Tabel 1 menunjukkan cuplikan hasil query yang didapatkan dengan jumlah data sebanyak 27 baris (Syarwani, 2020).

Tabel 1. Hasil Query

Politikus	Partai
Prabowo Subianto	Partai Gerindra
Fadli Zon	Partai Gerindra
Suhardi	Partai Gerindra
Marzuki Alie	Partai Demokrat
Alex Noerdin	Partai Golongan Karya
Haeny Relawati	Partai Golongan Karya
Rini Widyastuti	
Ason Sukasa	
Syahrial Oesman	PDI Perjuangan
Hatta Rajasa	Partai Amanat Nasional
Agus Hamdani	Partai Persatuan Pembangunan

3.2. Pemodelan Graph

Graph model menghubungkan antara seorang politikus dengan partai tempat dia bekerja atau pernah bekerja seperti yang direpresentasikan oleh gambar 4. Relasi yang digunakan adalah pekerjaan.



Gambar 3. Node dan Relasi Graf

3.3. Menjalankan Node2Vec

Parameter yang digunakan pada penggunaan Node2Vec adalah sebagai berikut. Dimensions berisi nilai 20, walk_length bernilai 16, num_walks bernilai 100, dan workers bernilai 2.

3.4. Menjalankan Algoritma Harmonic Centrality

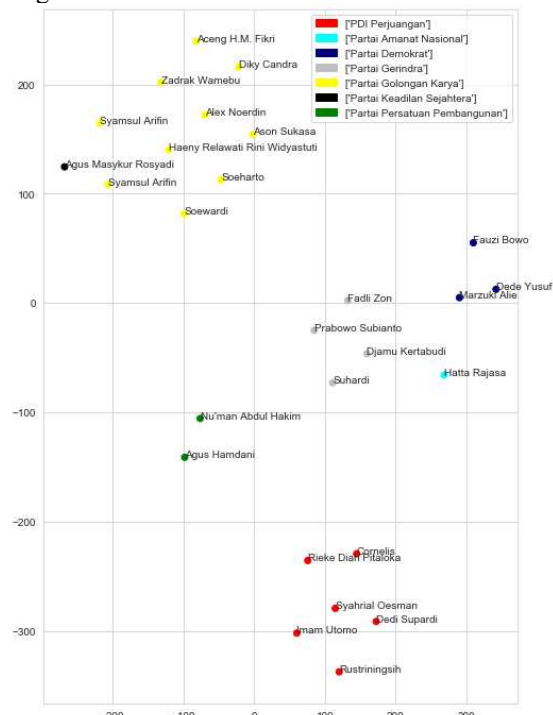
Parameter yang digunakan dalam menjalankan algoritma harmonic centrality yaitu undirected graph dan unweighted graph.

3.5. Visualisasi

Untuk visualisasi digunakan *tools* t-SNE dan Matplotlib. t-SNE dapat memvisualisasikan *high-dimensional* data dengan memberikan masing-masing datapoin sebuah lokasi pada *two* atau *three-dimensional map*. Hasil yang dihasilkan t-SNE memiliki visualisasi lebih baik apabila dibandingkan dengan teknik serupa lainnya karena hasilnya yang mengurangi kecenderungan untuk menumpuk datapoin ditengah (van der Maaten & Hinton, 2008). Matplotlib adalah *library* pada bahasa Python untuk membuat berbagai macam jenis visualisasi (Matplotlib, 2020).

4. Hasil dan Pembahasan

Dari data csv yang didapatkan dilakukan perubahan dari bentuk graf kedalam bentuk vektor dengan Node2Vec.

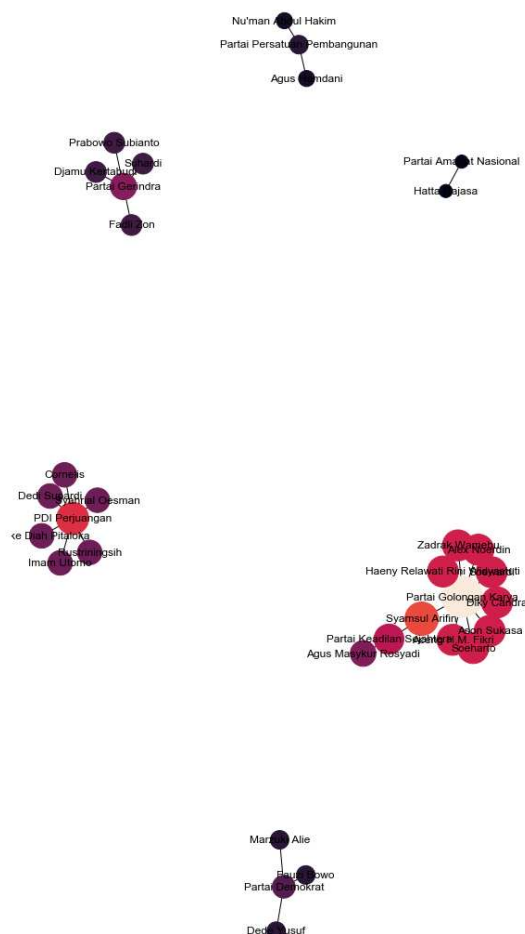


Gambar 4. Visualisasi t-SNE

Gambar 4 merupakan visualisasi t-SNE yang menunjukkan daftar politikus dan dikelompokkan berdasarkan hubungan kedekatan dengan partai mereka. Perbedaan warna pada politikus yang direpresentasikan sebagai datapoin pada dua dimensi menunjukkan politikus tersebut memiliki kelompok partai yang berbeda. Politikus dengan partai yang sama akan terletak berdekatan dan berwarna sama.

Dengan t-SNE, data yang direpresentasikan berupa vektor dengan dimensi tinggi dapat dikelompokkan dalam bentuk visualisasi dua dimensi yang mudah dipahami. Apabila dilihat pada gambar 4 terlihat bahwa kelompok partai dengan datapoin paling banyak yaitu Partai Golongan

Karya dan apabila diperhatikan kelompok Partai Golongan Karya berdekatan dengan kelompok Partai Keadilan Sejahtera dikarenakan kedua kelompok tersebut memiliki kemiripan yang sama dan masih terhubung melalui node politikus Syamsul Arifin yang pernah terafiliasi dengan kedua partai tersebut.



Gambar 5. Hasil Visualisasi Algoritma Harmonic Centrality

Gambar 5 merupakan visualisasi algoritma harmonic centrality dari networkx. Visualisasi dilakukan dengan parameter berupa ukuran dan warna yang menyesuaikan dengan nilai harmonic pada suatu node. Semakin besar nilai centrality yang dihasilkan maka akan semakin besar ukuran node dan warna node tersebut akan semakin cerah.

Dari visualisasi pada gambar 5 dapat dilihat Partai Golongan Karya memiliki warna dan ukuran yang mencolok apabila dibandingkan dengan node lainnya. Tingginya nilai centrality yang dihasilkan menunjukkan seberapa penting keberadaan node Partai Golongan Karya yang memiliki banyak relasi hubungan dengan node yang lain. Visualisasi tersebut

menunjukkan bahwa Partai Golongan Karya merupakan partai yang populer dan memiliki pengaruh besar dengan banyaknya politikus yang terhubung dengan node tersebut berdasarkan data dari DBpedia Indonesian pada halaman politikus.

Besarnya pengaruh yang dimiliki Partai Golongan Karya dengan politikus yang banyak terafiliasi dengan mereka dan partai lainnya. Hal ini menunjukkan Partai Golongan Karya memiliki kekuatan yang besar untuk menyebarkan pesan politik, serta keputusan politiknya dapat mempengaruhi partai-partai lainnya. Sebaliknya partai-partai dengan nilai centrality yang rendah seperti Partai Amanat Nasional yang hanya terhubung dengan satu node saja, memiliki pengaruh yang kecil dan keputusan politiknya tidak banyak berpengaruh terhadap partai dan politikus lainnya.

Berikut lima node dengan nilai centrality terbesar yang ditunjukkan pada tabel 2.

Tabel 2. Nilai Centrality

No.	Node	Nilai Centrality
1.	Partai Golongan Karya	9.84
2.	Syamsul Arifin	6.50
3.	PDI Perjuangan	6.00
4.	Alex Noerdin	5.58
5.	Haeny Relawati Rini Widyastuti	5.58

Graf yang direpresentasikan oleh gambar 4 dan 5 memiliki network density sebesar 0.05113636363636364. Hal ini disebabkan oleh banyaknya varian partai yang ada namun dengan sedikit orang yang berada pada partai tersebut.

Hasil pengolahan dengan algoritma harmonic centrality menunjukkan terdapat beberapa node yang memiliki pengaruh besar (Tabel 2) dan kecil. Node yang memiliki pengaruh besar diantaranya adalah Partai Golongan Karya, Syamsul Arifin, dan PDI Perjuangan. Asumsi besar kecilnya pengaruh didasari oleh nilai centrality yang dihasilkan oleh algoritma. Semakin besar nilai maka semakin besar pula pengaruhnya, begitu juga sebaliknya. Nilai centrality menunjukkan seberapa banyak node tersebut terhubung dengan node yang lainnya sehingga keberadaan mereka dapat mempengaruhi partai atau politikus yang lain. Sedangkan berikut beberapa node yang memiliki pengaruh yang kecil antara lain Hatta Rajasa, Partai Amanat Nasional, dan Agus Hamdani.

5. Kesimpulan

Dengan pendekatan berupa graf menggunakan algoritma harmonic centrality, data politikus yang didapat dari DBpedia Indonesian dapat dilakukan analisa terkait dengan partai/politikus yang memiliki kekuatan besar untuk

mempengaruhi partai/politikus lain. Hasil yang didapat menunjukkan terdapat partai/politikus dengan nilai sentralitas tertinggi yang terhubung oleh banyak node lainnya yaitu Partai Golongan Karya, Syamsul Arifin dan PDI Perjuangan. Partai Golongan Karya dan PDI Perjuangan memiliki kekuatan yang besar untuk menyebarkan pesan politik, serta keputusan politiknya dapat mempengaruhi dan diikuti oleh partai yang lain. Kondisi sistem politik Indonesia selama ini banyak dipengaruhi kedua partai tersebut. Banyak politikus yang menduduki kursi pemerintahan berasal dari kedua partai tersebut. Data memiliki banyak varian partai dengan sedikit orang yang berada pada partai tersebut, hal ini menyebabkan nilai network density menjadi cukup rendah dengan nilai 0.05113636363636364.

Daftar Pustaka:

- Boldi, P. & Vigna, S., 2013. *In-Core Computation of Geometric Centralities with HyperBall: A Hundred Billion Nodes and Beyond*. Dallas, 2013 IEEE 13th International Conference on Data Mining Workshops.
- Boldi, P. & Vigna, S., 2014. Axioms for Centrality. *Internet Mathematics*, 10(3-4), pp. 222-262.
- Grover, A. & Leskovec, J., 2016. *node2vec: Scalable Feature Learning for Networks*. New York, Association for Computing Machinery.
- Kadir, A., 2014. Dinamika Partai Politik di Indonesia. *Sosiohumaniora*, 16(2), pp. 132-136.
- Leskovec, J., 2016. *node2vec*. [Online] Available at: <https://snap.stanford.edu/node2vec/> [Accessed 6 Mei 2020].
- Marchiori, M. & Latora, V., 2000. Harmony in the Small-World. *Physica A: Statistical Mechanics and its Applications*, 285(3-4), pp. 539-546.
- Matplotlib, 2020. *Matplotlib: Python plotting*. [Online] Available at: <https://matplotlib.org/> [Accessed 7 Mei 2020].
- Putman, K. L., Boekhout, H. D. & Takes, F. W., 2019. *Fast Incremental Computation of Harmonic Closeness Centrality in Directed Weighted Networks*. Vancouver, 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
- R., P., V., V., Sharma, T. & Prakash, G. B., 2019. Predicting Harmonic Centrality In Adv For Wireless Sensor Networks Using Machine Learning. *International Journal of Scientific & Technology Research*, 8(10), pp. 804-808.
- Rochat, Y., 2009. *Closeness Centrality Extended To Unconnected Graphs : The Harmonic Centrality Index*. Zürich, s.n.
- Sugiarta, A. I., Dedy, S. & Negara, E. S., 2018. *Analisis Sentralitas Aktor pada Struktur Jaringan Politik dengan Menggunakan Metode Social Network Analysis (SNA) : Studi Kasus Group Facebook Lembaga Survei Sosial Media*. Palembang, Seminar Nasional Teknologi Informasi dan Komunikasi.
- Syarwani, A. R., 2020. *arifqys/data-politikus: Update release* | Zenodo. [Online] Available at: <https://zenodo.org/record/3838817> [Accessed 22 Mei 2020].
- van der Maaten, L. & Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, Volume 9, pp. 2579-2605.
- W3C, 2013. *SPARQL 1.1 Overview*. [Online] Available at: <https://www.w3.org/TR/sparql11-overview/> [Accessed 4 Mei 2020].

